IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): H. Chen et al.
Docket No.: YOR920030422US1
Serial No.: 10/693,238
Filing Date: October 23, 2003
Group: 2157
Examiner: El Hadji Malick Sall

Title: Methods and Systems for Dynamically Configurable Load Balancing

---

APPEAL BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Applicants (hereinafter referred to as "Appellants") hereby appeal the final rejection of claims 1-37 of the above-identified application.

REAL PARTY IN INTEREST

The present application is assigned to International Business Machines Corporation, as evidenced by an assignment recorded February 3, 2004 in the U.S. Patent and Trademark Office at Reel 14303, Frame 154. The assignee, International Business Machines Corporation, is the real party in interest.

RELATED APPEALS AND INTERFERENCES

There are no known related appeals or interferences.

STATUS OF CLAIMS

Claims 1-35 and 37 stand finally rejected under 35 U.S.C. §102(e). Claim 36 stands finally rejected under 35 U.S.C. §103(a). Claims 1-37 are appealed.

STATUS OF AMENDMENTS

There has been no amendment filed subsequent to the final rejection.

SUMMARY OF CLAIMED SUBJECT MATTER

Independent claim 1 is directed to a method of serving data to a plurality of clients in a client-server environment. The method comprises the steps of providing a plurality of versions of data in which at least two versions have different overheads associated therewith; assigning individual clients to one of a plurality of quality-of-service classes; and satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

In an illustrative embodiment described in the present specification at, for example, page 5, lines 16-25, with reference to FIG. 1, a method of serving data to a plurality of clients in a client-server environment comprises the steps of providing a plurality of versions of data in which at least two versions have different overheads associated therewith (e.g., high overhead version 15-1 and low overhead version 16-1 of content source 14-1 in FIG. 1); assigning individual clients to one of a plurality of quality-of-service classes (e.g., high QoS class client 11-H and low QoS class client 11-L in FIG. 1). As described in the present specification at, for example, page 10, line 24, to page 11, line 12, with reference to FIG. 4, the method also includes a step of satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

Independent claim 17 is directed to an apparatus for serving data to a plurality of clients in a client-server environment, comprising a memory, and at least one processor coupled to the memory. The at least one processor is operative to: (i) provide a plurality of versions of data in which at least two versions have different overheads associated therewith; (ii) assign individual clients to one of a

plurality of quality-of-service classes; and (iii) satisfy requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

In an illustrative embodiment described in the present specification at, for example, page 24, line 24, to page 25, line 22, with reference to FIG. 10, an apparatus (e.g., computer system 100 in FIG. 10) for serving data to a plurality of clients in a client-server environment comprises a memory (e.g., memory 104 in FIG. 10) and at least one processor (e.g., processor 102 in FIG. 10) coupled to the memory. As described in the present specification at, for example, page 5, lines 16-25, with reference to FIG. 1, the at least one processor is operative to provide a plurality of versions of data in which at least two versions have different overheads associated therewith (e.g., high overhead version 15-1 and low overhead version 16-1 of content source 14-1 in FIG. 1); and assign individual clients to one of a plurality of quality-of-service classes (e.g., high QoS class client 11-H and low QoS class client 11-L in FIG. 1). As described in the present specification at, for example, page 10, line 24, to page 11, line 12, with reference to FIG. 4, the at least one processor is further operative to satisfy requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

Independent claim 29 is directed to an article of manufacture for use in serving data to a plurality of clients in a client-server environment, comprising a machine readable medium containing one or more programs. When executed, the one or more programs implement the steps of: providing a plurality of versions of data in which at least two versions have different overheads associated therewith; assigning individual clients to one of a plurality of quality-of-service classes; and satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

In an illustrative embodiment described in the present specification at, for example, page 26, lines 4-7, an article of manufacture for use in serving data to a plurality of clients in a client-server environment, comprising a machine readable medium containing one or more programs. As described in the present specification at, for example, page 5, lines 16-25, with reference to FIG. 1, the one or more programs, when executed, implement the steps of providing a plurality of versions of

data in which at least two versions have different overheads associated therewith (e.g., high overhead version 15-1 and low overhead version 16-1 of content source 14-1 in FIG. 1); and assigning individual clients to one of a plurality of quality-of-service classes (e.g., high QoS class client 11-H and low QoS class client 11-L in FIG. 1). As described in the present specification at, for example, page 10, line 24, to page 11, line 12, with reference to FIG. 4, the one or more programs also implement the step of satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

Independent claim 30 is directed to a system, comprising: a plurality of clients, each client belonging to a quality-of-service class; a load balancer for sending requests from clients to at least one back-end server; and at least one back-end server for providing a plurality of versions of different objects in which at least two versions of an object have different overheads associated therewith.

In an illustrative embodiment described in the present specification at, for example, page 5, lines 16-25, with reference to FIG. 1, and page 6, line 22, to page 7, line 26, with reference to FIG. 2, a system comprises a plurality of clients, each client belonging to a quality-of-service class (e.g., high QoS class client 11-H and low QoS class client 11-L in FIG. 1); a load balancer (e.g., routing intermediary(s) 22 in FIG. 2) for sending requests from clients (e.g., received via front end servers 21 in FIG. 2) to at least one back-end server (e.g., high-end server(s) 23 and low-end server(s) 24 in FIG. 2); and at least one back-end server for providing a plurality of versions of different objects in which at least two versions of an object have different overheads associated therewith (e.g., high overhead version 15-1 and low overhead version 16-1 of content source 14-1 in FIG. 1).

Independent claim 31 is directed to a method of providing a data serving service, comprising the step of: a service provider: (i) providing a plurality of versions of data in which at least two versions have different overheads associated therewith; (ii) assigning individual clients to one of a plurality of quality-of-service classes; and (iii) satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

As described in the present specification at, for example, page 5, lines 16-25, with reference to FIG. 1, a method of providing a data serving service, comprises the step of: a service provider providing a plurality of versions of data in which at least two versions have different overheads associated therewith (e.g., high overhead version 15-1 and low overhead version 16-1 of content source 14-1 in FIG. 1); and assigning individual clients to one of a plurality of quality-of-service classes (e.g., high QoS class client 11-H and low QoS class client 11-L in FIG. 1). As described in the present specification at, for example, page 10, line 24, to page 11, line 12, with reference to FIG. 4, the service provider satisfies requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

Independent claim 37 is directed to a method of serving data to a plurality of clients. The method comprises the steps of: establishing at least two quality-of-service classes; and satisfying requests so that a client belonging to one quality-of-service class is served with a data version having one overhead associated therewith, while a client belonging to another quality-of-service class is served with a data version having another overhead associated therewith.

In an illustrative embodiment described in the present specification at, for example, page 5, lines 16-25, with reference to FIG. 1, and page 10, line 24, to page 11, line 12, with reference to FIG. 4, a method of serving data to a plurality of clients comprises the steps of: establishing at least two quality-of-service classes; and satisfying requests so that a client belonging to one quality-of-service class (e.g., high QoS class client 11-H in FIG. 1) is served with a data version having one overhead associated therewith (e.g., high overhead version 15-1 in FIG. 1), while a client belonging to another quality-of-service class (e.g., low QoS class client 11-L in FIG. 1) is served with a data version having another overhead associated therewith (e.g., low QoS class client 11-L in FIG. 1).

As described in the present specification at, for example, page 4, lines 1-5, illustrative embodiments of the present invention may advantageously provide efficient load balancing to back-end servers, as well as quality-of-service routing to provide some clients with better quality-of-service than others. Load balancing may follow policies that can be dynamically modified (reconfigured) without having to recompile the application or the intermediary code.

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

(I) Claims 1-35 and 37 are rejected under 35 U.S.C. §102(e) as being anticipated by U.S. Patent No. 7,020,697 to Goodman et al. (hereinafter "Goodman").

(II) Claim 36 is rejected under 35 U.S.C. §103(a) as being unpatentable over Goodman in view of U.S. Patent Application Publication No. 2004/0003080 to Huff (hereinafter "Huff").

ARGUMENT

Appellants incorporate by reference herein the disclosures of their previous responses filed in the present application, namely the responses dated September 19, 2007 and January 31, 2008.

(I) Claims 1-35 and 37 are rejected under 35 U.S.C. §102(e) as being anticipated by U.S. Patent No. 7,020,697 to Goodman et al. (hereinafter "Goodman").

With regard to the §102(e) rejection of claims 1-35 and 37, Appellants initially note that MPEP §2131 specifies that a given claim is anticipated "only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference," citing Verdegaal Bros. v. Union Oil Co. of California, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). Moreover, MPEP §2131 indicates that the cited reference must show the "identical invention . . . in as complete detail as is contained in the . . . claim," citing Richardson v. Suzuki Motor Co., 868 F.2d 1226, 1236, 9 USPQ2d 1913, 1920 (Fed. Cir. 1989). Appellants respectfully traverse the §102(e) rejection on the ground that the Goodman reference fails to teach or suggest each and every limitation of claims 1-35 and 37 as alleged.

Independent claim 1 recites a method of serving data to a plurality of clients in a client-server environment, comprising the steps of: providing a plurality of versions of data in which at least two versions have different overheads associated therewith; assigning individual clients to one of a plurality of quality-of-service classes; and satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve. Independent claims 17, 29-31 and 37 recite certain similar limitations, as well as other limitations. Advantageously, the claimed invention provides that clients belonging to higher quality

of service classes may be given preferential access to higher quality content (i.e., in many cases, higher quality content requires more overhead to serve).

Appellants respectfully point out that Goodman has absolutely nothing to do with serving data to a plurality of clients wherein requests are satisfied so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve (e.g., higher quality content).

The final Office Action cites column 69, lines 15-18; column 99, lines 52-61; and column 25, 3-21, in rejecting the steps of the independent claims; however, no where do these portions or any portions of Goodman teach or suggest the recited elements of the independent claims.

For instance, in rejecting the claimed step of providing a plurality of versions of data in which at least two versions have different overheads associated therewith, the final Office Action cites column 69, lines 15-18:

> . . . data management tools provide backup and restore facilities for data, and also provide configuration management for multiple versions of data, maintaining consistency among versions of test data.

However, Goodman clearly does not indicate that such multiple versions have different overheads associated with them.

In rejecting the claimed step of assigning individual clients to one of a plurality of quality-of-service classes, the final Office Action cites column 99, lines 52-61:

> The quality of service services 244 may also use data prioritization to improve network performance. While not an example of end-to-end QoS, various network components can be configured to prioritize their handling of specified types of traffic. For example, routers can be configured to handle legacy mainframe traffic (SNA) in front of other traffic (e.g., TCP/IP). A similar technique is the use of prioritized circuits within Frame Relay, in which the Frame Relay network vendor assigns different priorities to different permanent virtual circuits.

However, it is clear that Goodman only discloses prioritized handling of entire classes of traffic. Goodman does not disclose assigning individual clients to different quality of service classes. It can also be understood that Goodman does not disclose quality of service classes either.

Lastly, in rejecting the claimed step of satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve, the final Office Action cites column 25, lines 3-21:

> The CIP then plans and manages improvement related activities such as: define explicit criteria for assigning priority; consider raising the priority of low-priority opportunities that can be completed quickly; maintain a mix of high-priority and sure successes to ensure the continued momentum of the continuous improvement program; define the opportunity selection process; identify the resource allocation process; define the scheduling process; identify how the effort will be monitored; identify the procedure for communicating results to the organization; establish a continuous improvement organization to support the process; prioritize and classify opportunities; select projects; allocate resources and scheduling; monitor effort; and, support a standard process improvement process across the project. While maintaining quality at a program level, the quality management team 510 must liaise with each of the organizational units within the development architecture 500 in order to monitor the quality management processes within these units.

However, Appellants are completely unclear as to how this portion of Goodman (or any portion of Goodman) discloses that requests are satisfied so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve. This is because Goodman clearly does not teach or suggest such limitations.

Similar deficiencies are evident in the use of Goodman to reject the other independent claims.

The Advisory Action dated April 30, 2008 merely reiterates the same rationale first raised in the final Office Action and challenged in Appellants' prior response.

Regarding the dependent claims of the present application, it is asserted that they are patentable over the cited references not only due to their dependence of respective ones of the above-mentioned independent claims, but also because such claims recite separately patentable subject matter.

Dependent claims 2 and 18 recite the overhead to serve a version is correlated with a quality of the version. For example, the high overhead version is typically of higher quality than the low overhead version (Specification, page 6, lines 7-8). Giving comprehensive training to help desk personnel in order to ensure the best possible level of service to the developers (Goodman, column 27, lines 14-16) does not teach or suggest correlating the overhead to serve a version with a quality of the version.

Dependent claims 3 and 19 recite the plurality of versions comprise images of different resolutions and clients belonging to a high quality-of-service class are given preferential access to higher resolution images. As noted above, Goodman at column 99, lines 52-61 refers to prioritized handling of entire classes of traffic, which does not teach or suggest giving clients belonging to a high quality-of-service class preferential access to higher resolution images.

Dependent claims 4 and 20 recite the quality of a version is correlated with a processing time required to create the version. Goodman, at column 26, lines 33-40 states specifications of service levels should be precise and the service must be measurable, which does not teach or suggest a correlation between the quality of a version and a processing time required to create the version.

Dependent claims 5 and 21 recite the overhead to serve a version is correlated with how current the version is. Although Goodman at column 22, lines 4-12 refers to completed documentation being broken down by versions, the relied-upon portion of Goodman does not teach or suggest the overhead to serve a version is correlated with how current the version is.

Dependent claims 6 and 22 recite in response to a system load exceeding a threshold, satisfying a higher percentage of requests from clients belonging to a lower quality-of-service class with a version requiring lower overhead to serve. Goodman at column 25, lines 3-21 considers raising the priority of low-priority opportunities that can be completed quickly. Assuming the low-priority opportunities that can be completed quickly are the recited clients belonging to a lower quality-of-service class with a version requiring lower overhead to serve, Goodman does not teach or suggest raising the priority of low-priority opportunities that can be completed quickly in response to a system load exceeding a threshold.

Regarding dependent claims 7 and 23, nowhere does FIG. 4 of Goodman disclose the server comprising multiple nodes and different nodes providing data versions requiring different overheads to serve.

With regard to claims 8-11 and 24-27, the final Office Action refers to Goodman at column 79, lines 35-59 and column 98, lines 46-58 as teaching or suggesting the limitations of claims 8-11 and 24-27. Goodman at column 79, lines 35-59, refers to the preferred database replication/synchronization services supporting one or more of three ownership models and four basic types of replication styles. Goodman at column 98, lines 46-58 refers to quality of service services 244, which are application designed to deliver a predetermined level of quality throughout the network for designated traffic by allocating dedicated bandwidth, prioritizing data traffic, and managing traffic flow. Although the relied-upon portions of Goodman refers to different types of network traffic (e.g., data, voice, video), having different quality of service requirements, Goodman does not teach or suggest a quality-of-service policy that specifies at least one of content quality and latency (as recited in claims 8 and 24), where one or more clients belonging to a premium service class are served with high content quality and low latency (as recited in claims 9 and 25), where one or more clients belonging to a medium service class are served with one of high content quality and low latency (as recited in claims 10 and 26), and wherein one or more clients belonging to a best-effort service class are served with unspecified content quality and latency (as recited in claims 10 and 26).

Dependent claims 32 and 33 recite similar limitations and are believed allowable for at least the reasons noted above with regard to claims 8-11 and 24-27.

Claims 12 and 28 recite a client request is routed using at least one of an identity of the client, a quality of content, a load on at least one server, a data distribution on at least one server, and a capacity of at least one server. Although Goodman at column 108, lines 30-44 refers to data dependent routing, nowhere does the relied-upon portion of Goodman teach or suggest routing using at least one of an identity of the client, a quality of content, a load on at least one server, a data distribution on at least one server, and a capacity of at least one server.

With regard to claims 13, 14 and 16, the final office Action refers to Goodman at column 99, lines 52-61 and column 112, lines 25-36 as disclosing the limitations of claims 13 and 14, and column 25, lines 3-21 as disclosing the limitation of claim 16. Nowhere do the relied-upon portions of Goodman teach or suggest a client being assigned to a quality-of-service class by program logic that is externalized from the server (as recited in claim 13). Although column 112, lines 25-36 refers to business process rules, the relied-upon portion does not teach or suggest the business process rules can be modified by nonexperts in information technology (as recited in claim 14 and 16).

With regard to claim 15, Goodman at column 25, lines 3-21 states that while maintaining quality at a program level, the quality management team 510 must liaise with each of the organizational units within the development architecture 500 in order to monitor the quality management processes within these units. Nowhere does Goodman state that the quality management team's program logic determined a policy utilized in the step of satisfying requests.

The Advisory Action fails to address any of the above remarks, which were first presented in Appellants' prior response.

(II) Claim 36 is rejected under 35 U.S.C. §103(a) as being unpatentable over Goodman in view of U.S. Patent Application Publication No. 2004/0003080 to Huff (hereinafter "Huff").

Claim 36 recites that the step of assigning individual clients to one of a plurality of quality-of-service classes is based on a client payment.

For at least the same reasons as given above in section I with respect to the deficiencies of Goodman, Appellants assert that dependent claim 36 is patentable over the combination of Goodman and Huff. Also, Huff fails to remedy any of the deficiencies of Goodman.

Further, Goodman and Huff fail to meet the standard for a proper obviousness combination as set forth in the Supreme Court's KSR decision. See KSR v. Teleflex, 127 S. Ct. 1727, 1741, 82 USPQ2d 1385, 1396 (2007), quoting In re Kahn, 441 F. 3d 977, 988, 78 USPQ2d 1329, 1336 (Fed. Cir. 2006) ("[R]ejections on obviousness grounds cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.")

The scant statement given at page 8 of the final Office Action ("[o]ne would be motivated to do so to allow identifying relative priorities of the clients (abstract)") is clearly deficient rationale to support a combination of Goodman and Huff to reject claim 36.

In view of the above, Appellants believe that claims 1-37 are in condition for allowance, and respectfully request withdrawal of the §102(e) and §103(a) rejections.

Respectfully submitted,

/William E. Lewis/

Date: June 30, 2008

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946

APPENDIX

1. A method of serving data to a plurality of clients in a client-server environment, comprising the steps of:

providing a plurality of versions of data in which at least two versions have different overheads associated therewith;

assigning individual clients to one of a plurality of quality-of-service classes; and

satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

2. The method of claim 1, wherein the overhead to serve a version is correlated with a quality of the version.

3. The method of claim 2, wherein the plurality of versions comprise images of different resolutions and clients belonging to a high quality-of-service class are given preferential access to higher resolution images.

4. The method of claim 2, wherein the quality of a version is correlated with a processing time required to create the version.

5. The method of claim 1, wherein the overhead to serve a version is correlated with how current the version is.

6. The method of claim 1, further comprising the step of:

in response to a system load exceeding a threshold, satisfying a higher percentage of requests from clients belonging to a lower quality-of-service class with a version requiring lower overhead to serve.

7. The method of claim 1, wherein the server comprises multiple nodes and different nodes provide data versions requiring different overheads to serve.

8. The method of claim 1, further comprising the step of implementing a quality-of-service policy that specifies at least one of content quality and latency.

9. The method of claim 8, wherein one or more clients belonging to a premium service class are served with high content quality and low latency.

10. The method of claim 8, wherein one or more clients belonging to a medium service class are served with one of high content quality and low latency.

11. The method of claim 8, wherein one or more clients belonging to a best-effort service class are served with unspecified content quality and latency.

12. The method of claim 1, wherein a client request is routed using at least one of an identity of the client, a quality of content, a load on at least one server, a data distribution on at least one server, and a capacity of at least one server.

13. The method of claim 1, wherein a client is assigned to a quality-of-service class by program logic that is externalized from the server.

14. The method of claim 13, wherein the externalized program logic comprises a set of business rules that can be modified by nonexperts in information technology.

15. The method of claim 1, further comprising the step of satisfying requests using a policy determined by program logic that is externalized from the server.

16. The method of claim 15, wherein the externalized program logic comprises a set of business rules that can be modified by nonexperts in information technology.

17. Apparatus for serving data to a plurality of clients in a client-server environment, comprising:

a memory, and

at least one processor coupled to the memory and operative to: (i) provide a plurality of versions of data in which at least two versions have different overheads associated therewith; (ii)

assign individual clients to one of a plurality of quality-of-service classes; and (iii) satisfy requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

18. The apparatus of claim 17, wherein the overhead to serve a version is correlated with a quality of the version.

19. The apparatus of claim 18, wherein the plurality of versions comprise images of different resolutions and clients belonging to a high quality-of-service class are given preferential access to higher resolution images.

20. The apparatus of claim 18, wherein the quality of a version is correlated with a processing time required to create the version.

21. The apparatus of claim 17, wherein the overhead to serve a version is correlated with how current the version is.

22. The apparatus of claim 17, wherein the at least one processor is further operative to, in response to a system load exceeding a threshold, satisfy a higher percentage of requests from clients belonging to a lower quality-of-service class with a version requiring lower overhead to serve.

23. The apparatus of claim 17, wherein the at least one processor comprises multiple nodes and different nodes provide data versions requiring different overheads to serve.

24. The apparatus of claim 17, wherein the at least one processor is further operative to implement a quality-of-service policy that specifies at least one of content quality and latency.

25. The apparatus of claim 24, wherein one or more clients belonging to a premium service class are served with high content quality and low latency.

26. The apparatus of claim 24, wherein one or more clients belonging to a medium service class are served with one of high content quality and low latency.

27. The apparatus of claim 24, wherein one or more clients belonging to a best-effort service class are served with unspecified content quality and latency.

28. The apparatus of claim 17, wherein a client request is routed using at least one of an identity of the client, a quality of content, a load on at least one server, a data distribution on at least one server, and a capacity of at least one server.

29. An article of manufacture for use in serving data to a plurality of clients in a client-server environment, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

providing a plurality of versions of data in which at least two versions have different overheads associated therewith;

assigning individual clients to one of a plurality of quality-of-service classes; and

satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

30. A system, comprising:

a plurality of clients, each client belonging to a quality-of-service class;

a load balancer for sending requests from clients to at least one back-end server; and

at least one back-end server for providing a plurality of versions of different objects in which at least two versions of an object have different overheads associated therewith.

31. A method of providing a data serving service, comprising the step of:

a service provider: (i) providing a plurality of versions of data in which at least two versions have different overheads associated therewith; (ii) assigning individual clients to one of a plurality of quality-of-service classes; and (iii) satisfying requests so that a client belonging to a high quality-of-service class is given preferential access to data versions which require higher overheads to serve.

32. The method of claim 31, wherein the data serving service comprises a quality-of-service policy specification.

33. The method of claim 32, wherein the quality-of-service policy specification comprises:

a plurality of subscriptions, each subscription being specified by content quality and service latency, wherein a limited premium service subscription is served with high content quality in low service latency, a medium service subscription is served with a high content quality or a low service latency, and an unlimited best-effort service subscription is served with unspecified content quality and latency.

34. The method of claim 31, wherein the service provider modifies data content and how the data content is served to clients in response to one or more changing conditions.

35. The method of claim 34, wherein one or more changing conditions comprises a source of a bottleneck.

36. The method of claim 31, wherein the step of assigning individual clients to one of a plurality of quality-of-service classes is based on a client payment.

37. A method of serving data to a plurality of clients, comprising the steps of:

establishing at least two quality-of-service classes; and

satisfying requests so that a client belonging to one quality-of-service class is served with a data version having one overhead associated therewith, while a client belonging to another quality-of-service class is served with a data version having another overhead associated therewith.

EVIDENCE APPENDIX

None.

RELATED PROCEEDINGS APPENDIX

None.